

```

1 ATGCTGGGGG CAGTGAAGG CCCAGGTGG AAGCAGGCGG AGGACATTAG
51 AGACATCTAC GACTTCCGAG ATGTTCTGGG CACGATCAAG CACCCCAACA
101 TTGTAGCCCT GGATGACATC TATGAGAGTG GGGGCCACCT CTACCTCATC
151 ATGCAGCTGG TGTGGGTGG GGAGCTCTTT GACCGTATTG TGGAAAAAGG
201 CTTCTACACG GAGCGGGACG CCAGCCGCTT CATCTTCCAG GTGCTGGATG
251 CTGTGAAATA CCTGCATGAC CTGGGCATTG TACACCGGGA TCTCAAGCCA
301 GAGAATCTGC TGTACTACAG CCTGGATGAA GACTCCAAAA TCATGATCTC
351 CGACTTTGGC CTCTCCAAGA TGGAGGACCC GGGCAGTGTG CTCTCCACCG
401 CCTGTGGAAC TCCGGGATAC GTGGCCCCTG AAGTCCTGGC CCAGAAGCCC
451 TACAGCAAGG CTGTGGATTG CTGGTCCATA GGTGTCATCG CCTACATCTT
501 GCTCTGCGGT TACCCTCCCT TCTATGACGA GAATGATGCC AAACCTTTTG
551 AACAGATTTT GAAGGCCGAG TACGAGTTTG ACTCTCCTTA CTGGGACGAC
601 ATCTCTGACT CTGCCAAGA TTTCATCCGG CACTTGATGG AGAAGGACCC
651 AGAGAAAAGA TTCACCTGTG AGCAGGCCTT GCAGCACCCA TGGATTGCAG
701 GAGATACAGC TCTAGATAAG AATATCCACC AGTCGGTGAG TGAGCAGATC
751 AAGAAGAACT TTGCCAAGAG CAAGTGGAAG CAAGCCTTCA ATGCCACGGC
801 TGTGGTGCGG CACATGAGGA AACTGCAGCT GGGCACCAGC CAGGAGGGGC
851 AGGGGCAGAC GGCAGCCAT GGGGAGCTGC TGACACCACT GGCTGGGGGG
901 CCGGCAGCTG GCTGTTGCTG TCGAGACTGC TGCCTGGAGC CGGGCACAGA
951 ACTGTCCCC ACCTGCCCC ACCAGCTCTA G (SEQ ID NO:1)

```

FEATURES:

Start Codon: 1

Stop Codon: 979

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 18000004983962 /altid=gi 4502553 /def=ref NP_003647.1 (NM...	661	0.0
CRA 18000004936440 /altid=gi 3122310 /def=sp Q63450 KCC1_RAT CA...	642	0.0
CRA 223000002652742 /altid=gi 15928726 /def=gb AAH14825.1 AAH14...	641	0.0
CRA 18000005144641 /altid=gi 3114436 /def=pdb 1A06 Calmodulin...	556	e-157
CRA 18000004932361 /altid=gi 406113 /def=gb AAA19670.1 (L24907...	548	e-155
CRA 117000066864297 /altid=gi 9966875 /def=ref NP_065130.1 (NM...	470	e-131
CRA 149000126089143 /altid=gi 14422219 /def=emb CAC41379.1 (AL...	398	e-109
CRA 224000007378166 /altid=gi 16755792 /def=gb AAL28100.1 AF428...	398	e-109
CRA 114000110934306 /altid=gi 14196445 /def=ref NP_065172.1 (N...	398	e-109
CRA 18000005191499 /altid=gi 4007153 /def=emb CAA19296.1 (AL02...	398	e-109

FIGURE 1A

1 MLGAVEGPRW KQAEIRDY DFRDVLGTIK HPNIVALDDI YESGGHLYLI
 51 MQLVSGGELF DRIVEKGFYT ERDASRLIFQ VLDVAVKYLHD LGIVHRDLKP
 101 ENLLYYSLDE DSKIMISDFG LSKMEDPGSV LSTACGTPGY VAPEVLAQKP
 151 YSKAVDCWSI GVIAYILLCG YPPFYDENDA KLFEQILKAE YEFDSPYWDD
 201 ISDSAKDFIR HLMKDPEKR FTCEQALQHP WIAGDTALDK NIHQSVSEI
 251 KKNFAKSKWK QAFNATAWR HMRKLQLGTS QEGGQTASH GELLTPVAGG
 301 PAAGCCCRDC CVEPGTELSL TLPQL (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

Prosite results:

PDOC00001 PS00001 ASN_GLYCOSYLATION

N-glycosylation site

264-267 NATA

PDOC00004 PS00004 CAMP_PHOSPHO_SITE

CAMP- and cGMP-dependent protein kinase phosphorylation site

219-222 KRFT

PDOC00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 3

1	28-30	TIK
2	70-72	TER
3	204-206	SAK

PDOC00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site

Number of matches: 9

1	55-58	SGGE
2	70-73	TERD
3	107-110	SLDE
4	122-125	SKME
5	204-207	SAKD
6	236-239	TALD
7	245-248	SVSE
8	279-282	TSQE
9	289-292	SHGE

PDOC00007 PS00007 TYR_PHOSPHO_SITE

Tyrosine kinase phosphorylation site

62-69 RIVEKGFY

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 3

1	128-133	GSVLST
2	283-288	GQGQTA
3	299-304	GGPAAG

PDOC00100 PS00108 PROTEIN_KINASE_ST

Serine/Threonine protein kinases active-site signature

93-105 IVHRDLKPENLLY

FIGURE 2A

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	127	147	0.883	Putative
2	155	175	1.515	Certain
3	288	308	0.649	Putative

BLAST Alignment to Top Hit:

>CRA|18000004983962 /altid=gi|4502553 /def=ref|NP_003647.1|
 (NM_003656) calcium/calmodulin-dependent protein kinase I
 [Homo sapiens] /org=Homo sapiens /taxon=9606 /div=PRI
 /dataset=nraa /length=370
 Length = 370

Score = 661 bits (1688), Expect = 0.0
 Identities = 326/370 (88%), Positives = 326/370 (88%), Gaps = 44/370 (11%)
 Frame = +3

Query: 126 MLGAVEGPRWKQAEDIRDIYDFRDVLGT----- 209
 MLGAVEGPRWKQAEDIRDIYDFRDVLGT
 Sbjct: 1 MLGAVEGPRWKQAEDIRDIYDFRDVLGTGAFSEVILAEDKRTQKLVAIKCIAKEALEGKE 60

Query: 210 -----IKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR 353
 IKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR
 Sbjct: 61 GSMENETIAVLHKIKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR 120

Query: 354 LIFQVLDAVKYLHDLGIVHRDLKPENLLYYSLDEDSKIMISDFGLSKMEDPGSVLSTACG 533
 LIFQVLDAVKYLHDLGIVHRDLKPENLLYYSLDEDSKIMISDFGLSKMEDPGSVLSTACG
 Sbjct: 121 LIFQVLDAVKYLHDLGIVHRDLKPENLLYYSLDEDSKIMISDFGLSKMEDPGSVLSTACG 180

Query: 534 TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAIEYFDSP 713
 TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAIEYFDSP
 Sbjct: 181 TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAIEYFDSP 240

Query: 714 YWDDISDSAKDFIRHLMKEDPEKRFTCEQALQHPWIAGDTALDKNIHQSVSEIQIKNFAK 893
 YWDDISDSAKDFIRHLMKEDPEKRFTCEQALQHPWIAGDTALDKNIHQSVSEIQIKNFAK
 Sbjct: 241 YWDDISDSAKDFIRHLMKEDPEKRFTCEQALQHPWIAGDTALDKNIHQSVSEIQIKNFAK 300

Query: 894 SKWKQAFNATAVVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT 1073
 SKWKQAFNATAVVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT
 Sbjct: 301 SKWKQAFNATAVVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT 360

Query: 1074 ELSPTLPHQL 1103
 ELSPTLPHQL
 Sbjct: 361 ELSPTLPHQL 370 (SEQ ID NO:4)

FIGURE 2B

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	271.6	1.1e-77	1
CE00022	CE00022 MAGUK_subfamily_d	119.7	3.1e-35	1
CE00359	E00359 bone_morphogenetic_protein_receptor	6.5	0.36	1
CE00031	CE00031 VEGFR	4.3	0.2	1
PF01496	V-type ATPase 116kDa subunit family	1.8	7.5	1
CE00292	CE00292 PTK_membrane_span	-89.8	0.0011	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-96.6	0.044	1
CE00291	CE00291 PTK_fgf_receptor	-123.4	0.11	1
CE00286	E00286 PTK_EGF_receptor	-151.4	0.095	1
CE00290	CE00290 PTK_Trk_family	-204.5	0.4	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-271.7	0.12	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF01496	1/1	85	95 ..	1	11 [.	1.8	7.5
CE00359	1/1	93	146 ..	272	327 ..	6.5	0.36
CE00031	1/1	79	163 ..	1053	1137 ..	4.3	0.2
CE00290	1/1	2	212 ..	1	282 []	-204.5	0.4
CE00292	1/1	3	227 ..	1	288 []	-89.8	0.0011
CE00291	1/1	1	230 [.	1	285 []	-123.4	0.11
CE00287	1/1	2	230 ..	1	260 []	-96.6	0.044
CE00286	1/1	1	230 [.	1	263 []	-151.4	0.095
CE00022	1/1	30	232 ..	75	283 ..	119.7	3.1e-35
PF00069	1/1	25	232 ..	43	278 .]	271.6	1.1e-77
CE00016	1/1	1	302 [.	1	433 []	-271.7	0.12

FIGURE 2C

```

1 AAACCGACCT TTGGCCTCTT GCCTGCCGTC CTAGTTGCAG GCTCTCTCCC
51 CTAACCTGGA CCCCAGCCAT CAAACTCTGG AGCCCCGCCA GTCACGTGAC
101 ACCTCGGTCC TTTTGGCCT GTTTCCTTCA GGATCCCGAT TTAACCTCCT
151 CCTCCCAAT TCCCTCTGCC CCAATACCT CTAGGCACCA CCACCCGCTC
201 TGAGGAGCAA GTGTCTGGGG CTGAAGCCTC AGCTCCATCT TGCAGAGGAA
251 CCGGGGCTC AGTCTTCCCA CTGTCAAGT GGGGCCACA CCCTGCGACC
301 ACCTCCACTC TCTTCATTGC CTAGTCTTGC CCGGTCTTC CCCACTCCCT
351 CACTCCCCCA TCCCCACCA GACTCCCGTG CAGTTCAGG GCCTGTTTCC
401 CTTAGGGCA CGGAGAAGGG AGACAGAGCC CTAAGGGAGG TCGCAGAACT
451 GGTCTGAAAG AAAATCCACC AGGCCACAGG GTGAGTTTGG CCGGCCTCTA
501 GCTTCAGACA GACGGGGTTC GAATCCTGCT TTGCTTCCGA CCACCCGCTG
551 ATTTGAAAAT CATCTCTCCG GGCCTCAGT GTCCCTCTG TGAATGGAC
601 CCCGCTTAAG ACCAAGGGCG GGAAGCGTCC AGCAGGAGAT CTCTGACCAG
651 AAGCAGGGAG ATGGCCTCCA CCCGTGCCCC TTCCCCAGCC TTGGAGCGGT
701 GCCTCGCTC CCAATCCCGG GTCCCTCCGC CGCAGGCTCC ACCTCCACTG
751 ACATCAGAGC CGCAGGCGGG CGGAGAGAGC CGCCGAGCCG AGCCGAGCCC
801 CAGCTCCAGC AAGAGCGCGG GCGGTGGCC CAGGCACGCA GCGGTGAGGA
851 CCGCGGCCAC AGCTCGGCGC CAACCACCGC GGGCCTCCA GCCAGCCCCG
901 CGGCGGGGCA GCCGCAGGTA CAGCCGGGCC CCCATCCCT GCACCCCTGG
951 GCGTGCCTG GGGGCGGTGG GAGCCCCTAG CCTCTGGGTA TCCTTCCCA
1001 AGGAGTGGCC ACTGGGCACT CTCCCGGGCG GGCTGGACCC TGAGGGGCG
1051 GGCTGGGCT TTCTCCACCT CTGTCCAGG CCCAGCAGGT GCCAGGCGGG
1101 CCTATGGGAC ACTGAGTGGG TAATAGAGAA GGGGGCCTGT GTGAGCGCCT
1151 TCAGCTGGGC CTGACTGGAA GGGCGTGGG ATTTGGAGGT ATCCATGGGG
1201 TGGGGGGGCT TGCGGAGTGT ACTGTCTTAG GACAGGCGTG TGGGTCAGAC
1251 ATGGGTGGAG GATCTGGGAA TCTGTGTGT TTTTGTCCA GAGGGGTGTC
1301 CACGTGTTTT GTGTGCTGGT ATTTGGCTCT CAGGGTCTTA AGTCAGAGTT
1351 AGGAGGGGGT GTACAATTGT GACTGAGGA TGTTTGGAGT TAGGTGTGTA
1401 AGGACTTGGG GTTTGGTTTG GAATACAGGA GCTTCCAGGG GATGGGGTAG
1451 AGGAGCTGGA GGGGTGAGGG TACGTCTGGT ATATGAGGGT GTGTGTGTGT
1501 GTGTCTGGGT GTCATCTTGT GTGGGTGCGG GTGGATGTGT GTTTTGGGGT
1551 GTAAGAGGGA GCTGGGTGAG GGATGTTTGG ATGGACAGGC AGGTGTTCCG
1601 GTGCAGGGCT GTCTGGGGCA CTGTGTGGTG TGGACATGTG TGCTGATGTC
1651 TGGGAGTACA TGTATGATCA GGTGTCACGG GATGTGGATA CAAGGCGTAC
1701 TGGATCTGGG AGGCAGGTGT TTGAGTTCAG GGCTGTGGAG GGGGCTTGGT
1751 GTGGCATGTC TGCTACAGGG ATGTGTGTGG ATCTGTGAGG GTTGTATTG
1801 GTAGGCCTCC ATGTGGGTTT CAGACTCTGC CTCTAGAGCT TACACTCGAG
1851 TCTCCTTTCC TAGAAGATTG TGCCCCTGGA TGGGTGGGCA GGGTCCCCTG
1901 GGAAAAAGGT CCTGTTCCAG GAGTGAATC TCACACCAGA GGCCCTAGTC
1951 AGGGCACCTT CTCCTCATTC TCCCTTAGAG AAAAAGAGAG AAGGAAAGTG
2001 CTCTCCCTGA GGTCACAAAG CATGCTGGGC TCTGTTTTGG CCTCATCTGT
2051 GGATGGGTTG GGAGGCTGTG TTCTCTGAAT GGGGCCATT CTGGCTTCAT
2101 ATTGGAAGTA CCAGCCAAGG CCATTCGATG GCCTTTGCCC TCAGCAAGCT
2151 TAGCTGGGGG CCCCAGGCCA GGTGTCATTA GGGCCTCTGG AGCCAGCCTC
2201 TACCTAACTC CAACCTCAGT CTCCCCATTC TTCATCTGAT AAATGGGAGA
2251 GAACTCCAC CCTCTCCTGC TGGATGAGAC AGACCTCAGC AGAGGAAGGG
2301 CCAGGCTGGA TAGGGTTAGA TGGGGCCAGG AAGGGACAGA GTGAGCAGGA
2351 CCATTTCTCA TGCTCCCGGG ACCCAGATGG GGAGTCAGGA GGGAGAGGTC
2401 TGGGGAGCTC CAGCTGTGGC TGTTGTTGCT GTGGTAACAG TGCAGAAAGA
2451 GCTATTTAAA AATGTGGCTG AGATGTTGCT GGAAGCCAG GCTGCTGGAA
2501 ACCTGATTTT GGAGAGGCCG GGGAGTCGGG GGAAGGAGGA GGGAAAGGAG
2551 ACACCCACGC AATCCCCAGG GTGGGGCGGG GACATCACTG GTTCTGGGGA
2601 CAGGGGGATC CTCCAGGCTT CTACCAGCTG CTCTGGGGGT TTATCTGTTG
2651 TACTGCCAGA AGTCAGGGTT TCCCTAGGTG CTTGGATTTG GATAGGGGGA

```

FIGURE 3A

2701 AAAGTGGGAA GAGAACTAGA ATAAATGAAT GAATGAATGC ATGACTTTGT
 2751 TAAATAAAGA ATTTTGCTGC CACTGTGAAA GGTTTTTCTC TAGGCATGAG
 2801 AATTTGCTGA ATGTTGAATA AACAAATGAA TGTTTGTTGA ATGATTTTGT
 2851 CAAATGGATG AATCAAGGAT GAATAAATGC AGGTTGAATG ACTGAATGGG
 2901 GCCTGCAGTA AATCCCAGA CAGAGGGCTG GGCTCTGCTG AGTCTCCTCC
 2951 TTCCATTCTC CTTACAGGAG CCCTGGCTGT GGTGCGGGGG CAGTGGGCCA
 3001 TGCTGGGGGC AGTGAAGGC CCCAGGTGGA AGCAGGCGGA GGACATTAGA
 3051 GACATCTACG ACTTCCGAGA TGTTCTGGGC ACGTGAGTCC AGGGCAGGAT
 3101 TGGGTGCTGG ATGGCTGAGG GAGGCTGAGT CCAGGGTGGG GCTTCCTCTG
 3151 GTCAATTAAT GCTTCCTGTT TCCCACAGCC CAGGCCCTGT GGCAGCACTA
 3201 TCTAGGGCCT AAAGTGTCCC CAGCTTTTCA CTTCTGGATG ACAGTGGGTG
 3251 GGACACGGGC TGCTCTCCA ATAGCCCTGG GTTCTTGAAG AGAAAGAAGT
 3301 CGAGAGAATG AAGGTGCCAG TCAGTCCATT TAACTTGCTG CCAAGAGCTA
 3351 AGTGTCTAG CCTAGGTTTG GGAAGTGAAG CTGGAGATGG CTCTGTTCTT
 3401 GGTGCTGGGA ATGCAGAAAT AACTCAAACC TGGTCTCTGC CCTTCAAGTT
 3451 GATCCCAGAC ATGTGCAAGA GACAGACCTA CAGAAAATGA CAACAGGGTG
 3501 TGTGCTGTGC TCCAATTAAG GTTGGGATTG AGGGCTTTGT GGAGCCCAGA
 3551 GAGAGCTGTG CCTTCTGCCT GGGGGAAGAC TTCCTGGAGA ATGGGGCATT
 3601 AGAGCTGGGG ACTGAAGGAT GGGTAGGTGT GCACTTGTCA GAGAGGAAGA
 3651 AGGACATTCC AGGCAGAAGG AATAGCATAA ACAAAGGCTT AGAGGCATGG
 3701 TTCTATGTGG AGAGAGGTAG AGTGTGATGG AGCTTAAAT CACAGGCTGG
 3751 GGGGAGAGTG GAAAAAGGGG CTGGAGATGA AAGTGGGACA GTTTGTGTAG
 3801 GGTTTTGGAA GCCAGGCCAG GGAGGCTGGA TATTGTCCA TAGGCCACCG
 3851 GGAGACACTT AAGACTTTTT GGCAGGTGTG CAATTCAGGA TAGTCACTCT
 3901 GGCCACAGCT TGGAGGGTAA ATTGGAGAGG GACAAGACTG GAAACCAAGT
 3951 ATGAGGTTAC TACAGTAACT AATTATCCCT GAGGATTGAA ATTTACCAC
 4001 GAGAGATGCT TTTCTTTGAC TTATGACTTC TTATTCTCCC AGAGAAAAGCA
 4051 AACAGATGTG GAAAGAATAC CCTAGCAAAT CCTCTTTAAT CAGTTAACTT
 4101 TAGTTAAATG AGTTTATTTG TTCCTTTTTA AGAACCTGTT CTAAACACT
 4151 GCTTCTTAAA GTTCAATGAG CATACAAATC ACCTGAGGAT TTTGTTAAAC
 4201 TGCAGATTGA TTTAGTAAAT CTGGGGCAGG GCCTAAAGTT TTGCATTTCT
 4251 TTTTTTTTTT TTTTTTTTGA CCCAGGATCC AAAGCAGTAG AGATTTTGCA
 4301 TTTCTAAAAA AGTTCCCGGG TGATGCTGAT GGTTCTTTAA GGTCTTAAAG
 4351 GGTGTTAAAT TAGCCATGAC TCGAATTAGC AGAAAAAGGG ATGAACCAAC
 4401 TGTACACATA ATCCAAAAGC CCAGGGGTAG ACCTCAGGCA TGGCTGGATC
 4451 CAGAGGGCCA CATAATGTTA TCAGGAAATA TATTTGGCCA TTTCTCAGGT
 4501 TGGACTTCCT TTGTGTTAAT TTCATTCCCA AGCAGGCTCT CCCCAGGTGG
 4551 TGGCAAAGAT GATCGCCATT AGCTCCAGGC TTACATCCTA CCAGCTCAAC
 4601 AGGAGACTCA TTCTCAAAGT GCTAGTAAGC TGGCTTGCA CACATGACCA
 4651 ATTACTGTGG CCAGGGGAGA GACTACTTTG ACTGGCCAGG CCTGGGTGAT
 4701 GTGACCATCT CTGGAGCCAG GGGATGGATG AGTGACTAGG GGAGGGTCAT
 4751 CCACGTCCCT GGTCCAGCAG TGGTCACAGA ACCCATAGGG AATGGAGGAG
 4801 AGGCTGGAGG GAAGCTGGGG TTCCAGTTCT TCACCTTGAG AATCCCTCT
 4851 CCCGATAGGG GGGCCTTCTC GGAGGTGATC CTGGCAGAAG ATAAGAGGAC
 4901 GCAGAAGCTG GTGGCCATCA AATGCATTGC CAAGGAGGCC CTGGAGGGCA
 4951 AGGAAGGCTG CATGGAGAAT GAGATTGCTG TCCTGCACAA GTGCGTGGGC
 5001 CACAGCCTTT CCCTGCCCCA AGCTGACCCT GCCTTGGCCC TCCCATCCTC
 5051 CTCCTTTCTT GCTTTGGACA AATCATTTAA ACTCTCTAAG CCTTAAATTG
 5101 CCCCTTTATA AAATGGGGAT CACAATTTCC ACTTGGCAGG GTTGTGGGGA
 5151 ACATCAGAAG TCCTTTATTT CAAGTGCCTG GCCTAACATG ACAGATGTGA
 5201 TGGAGGTGCC AGTGCTTAGT CACAGGGGTT TAACTGTTCA ATCAGGTGTA
 5251 AAGATCCATC CTGAACATGG CTTGGACCCA CATATCTCAG TTGGTGTGTG
 5301 CTCTGGACCT ACCTCAAGTT CCCCTCACAT ATTAACCACT CTCAGCAAGT
 5351 TTAAAAATGA CTGTCTGCTG ACCCCAGAC TAAATCCACA ACCAACTGGT

FIGURE 3B

5401 CTATGAATTG CTCATGCTGA TATGAAACCT CCTGTCCTCA CTGGAAAACT
 5451 TACAGAGAAT CACTTCCAAT CTCTCCCCTG AGCTTCCAAC CACCCTGGGC
 5501 AGATAATTTT TTTTTTTTTT TTGAGATGGA GTCTCACTCT GTTGCCCCGG
 5551 CTGGAGTGCA GTGACGCAAT CTTGGCTCAC TGCAACCTCT GCCTCTTGGG
 5601 TTCAAGCAAT TCTCTTGCTT CAGCCTCCCT AGTAGCTGGG ATTACAGGCA
 5651 CCTGCCACCA CGCCCGGCTA ATTTTGTAT TTTTAGTAGA GATGGGGTTT
 5701 CGCCATGTTG GCCAGGCTGG TCTCGAACTC CTGACCTCAG GTGATCCACC
 5751 CGCCTCGGCC TCCCAAAGTG CTAGGCATGA GCCACCACAC CCAACTCCTG
 5801 GCAGAGCATT TCTAATAAGA CCCAGAGAGG ACAGGGATTT GTATACAGTC
 5851 ACATGGCAAG TTTGTGGCAG AGCTGAGCCT TCCTCATCAT CAAGATCAAT
 5901 TATCGCCTGA CCAACACGGA GAAACCCTGT CTCTACTAAA AATACAAAAT
 5951 TAGCCAGGCG TGGTGGCACA TGCCTGTAAT CCCAGCTACT TGGGAGGCTG
 6001 AGGCAGGAGA ATTGCTTAAA CCCGAGAGGT GGAGGTTGCG GTGAGCCGAG
 6051 ATCACACCGT GCATTACACT CCAGCCTAGG CAACAAGAGC AAAACTCCAT
 6101 CTCAAAAAAA AAAAAAAAC AAAAAAAA CAAAAACGCC AGGCGCAGTG
 6151 GCTCACGCCT GTAATCCCAG CACTTTGAGA GGCTGAAGTG GGCAGATCAC
 6201 CTGAGGTGGG GAGTTCCAAA CCAGCCTGAC CAACATGGAG AAACCTCCGC
 6251 TCTACTAAAA ATACAAAATT AGCTGGACAT GGTGGCGCAT GCCTGTAATC
 6301 CCAGCTACTT GAGAGGCTGA GAAAGAAGAA TCACTTGAAC CCAGGAGGCA
 6351 GAAATTGTGA TGAGCCAAGA TCATGCCATT GCACTCCAGC CTGGGCAACA
 6401 CTCCAGCCTG AGCAACAAGA GTAAAACTCC GTCTCAAAAA AAGAAAAAAA
 6451 AAATCAATTA CCATTATTGT TTCATTATG AGTATTTACC GTGTGCCAGG
 6501 CACTGTGCCA AGCACCTTAC CTGCATTATC TCACATGATC CTCCTCCAA
 6551 CTCTTTGAGG GAAGTACTAC CATTGGCTTC ATTTTATAGA TGAAGAACT
 6601 GAGGTTTACA GAGGTTACAT TAAATCTAGC ACCTACCCTG TACCAGGTGC
 6651 TGGAGGAACA GTGGCAAGCA AGACAAAGCC TCTGGATTCG GGGAGCTTAT
 6701 GTCTGGTGGG GGAGGCTGAC AAACATGTAA ACACAGAAAA CTATATATAT
 6751 ATATTTTTTT TGAGATGGAG TTTTGCTCTT GTTGCCAGG CTGGAGTGTA
 6801 ATGGCATGAT CTCGACTCAC TGCAACCTCC GTTTCACAGG TTAAAGCAAT
 6851 TCTCCTGCCT CAGCCTCACA GATAGCTGGG ATTACAGGCA TGTGCCACCA
 6901 TGCCTGGCTA ATTTTGTAT TTTTAGTAGA GATGGGTTT CGCCATGTTG
 6951 GCCAGGCTGG TCTCGAACTC CTGACCTCAA GTGATCCGCC TGCCTTGGCC
 7001 TCCCAAAGTG CTGGGATTAC AGGTGTGAGT CTCTGTGCCT AGCCAGAAAA
 7051 CTCTTAAGAG GTATGTATCA GGCTGGGTGC AGTGGCTCAC TGGTGAAAAG
 7101 ATCTGCACCC AAATAGCATG TGACGGGCAG GATTTGGACC CAGGTCTGTG
 7151 TATGCCAGAG CCCAGTGTTT ATCCCTCTGC TCTCTACCT TCCAAAAAAT
 7201 GGTAATAAAC CATGGTAAGC TAGCTTTTCC CTTTGGGGAC GAGATCCTTG
 7251 GTTTGTCTTA CCCAGGTATG TAGGCAGTGG TCGGGGGTTG GGGGTGGCTG
 7301 AGCTGTCTTG AGCTCTAAAC CGCTGTTTTT TTTTTTTTTT TTTTGAGACA
 7351 GGGTCTTACT CTGTTGCCCC GGCTGGAGTG CAGTGGCTAG TCACAGGTGC
 7401 AATCATAACA GACTGCAGCT TTGAACTGCT GGGGCCAAGT GATCCTCCTG
 7451 CCTCAGCCTC CCAAGTTCCC AAGTAGCTTG GACTACAGGT GCACACCGCC
 7501 ATGCCTGGCT AAACCACCTC ATTTCTCCTT TCAGGATCAA GCACCCCAAC
 7551 ATTGTAGCCC TGGATGACAT CTATGAGAGT GGGGGCCACC TCTACCTCAT
 7601 CATGCAGCTG TGAGTGGCCC AACCTCTGCC CTGCCCCCAC ACCTCTCCCA
 7651 GCTGTCCCAA CCCTCTTTGC CAGACTGCCC TATCCCCTGC TGCAGGGTGT
 7701 CGGGTGGGGA GCTCTTTGAC CGTATTGTGG AAAAAGGCTT CTACACGGAG
 7751 CGGGACGCCA GCCGCCTCAT CTTCCAGGTG CTGGATGCTG TGAAATACCT
 7801 GCATGACCTG GGCATTGTAC ACCGGGATCT CAAGGTGGGG CTCAAGGGGG
 7851 TGTGGTGAGC TAGGGTACCC AGGGGTGGGG CCTTTGCAAA CCCCAACTG
 7901 TCTGACCTTG GGCAACTTTC ACCCCCTCAC TGAGCCTTGG ATTTCCATCT
 7951 ACAAAGTGGA TCTTGTAACC TTAAACTGCT CTCCTCCCAT TCTAGTCCAG
 8001 ATACTCAAAG GAACACGAGT GAATTGTGTG GCATTTTATC CAAACAACAT
 8051 TTTGTCTTTT TCTGATTAAA AAAAAAAA TCTGGCCAGA CAGGATGGCT

FIGURE 3C


```

8101 CACGCCTGTA ATCCCAGCAC TTTAGGAGGC AGAGACGGGT GGATCACCTG
8151 AGGTCAGTTC GAGACCAGCT TGGCAAAACC CTGTCTCTAC CAAAAATACA
8201 AAAATTAGCC CGGCGTGGTG GCAGATGCCT GTAATCCCAG CTA TAGGGA
8251 GGCTGAGGCA GCGAATCAC TTGGACCCGG GAGGCAGAGG TTGCAGCAAG
8301 CTGAGATTGT GCCATTGCAC GCCAGCCTGG GCGACAGAGC GAGCCTGGAC
8351 GACAGAGCGA GACTCCATGT CAAAAAAAT AAAATAAAAA CAAAAAATCC
8401 TATTCCCCTT CTGTAGAAAA CTTGGATGGG ACAGCAAAAC ATAAAGAAAA
8451 AAGCCAGAAA TCCCCGAAAT CCTACTCCTC GGAAATAGCG ACGGGGCTCA
8501 CATTTAGCAG TACATCTCAA TCCGTTCTAG GAGAAGGGCA CTTGGGGTGT
8551 GACATGCCTG GTTTTGAATT CTGGCTCTGC TACTGCCTAA CTGTGGGTTC
8601 TTGGGTGAGT CACTTTGCCT CCAAAGGCAT CAGTTTCCTC ATCTGTTAGG
8651 TGAGATTATA CAGACTGGCC TAGCAGGGAA GCAGTGAGGA TGGCATTAAA
8701 TCAAGCACTA ATCCAGGGTC TGGCATAAAA TAGGCATTCA AACATTCCTT
8751 TAGGGCTTTA CAGTGCACAC CTGAGGTTTA GAGACAGTTC CCCCCACAC
8801 CCTCTTGAGC CTTGTCCCTC CTGGAATTTT TGGCCTTCTT GAGAGCTTCC
8851 TTGATTTTCT TATGACAGCC ATGAAGCCAC AGTGGCTTTT GGGGATCCAT
8901 TATTTCTCAG AAGGTGCTTG GAGCGGCAGA AGGTTCTACC AGCCTCTAAC
8951 CATCTCTGAT TGCCCCCTCT CTCCCTCCTC GCCCTTCAAG CCAGAGAATC
9001 TGCTGTACTA CAGCCTGGAT GAAGACTCCA AAATCATGAT CTCCGACTTT
9051 GGCTCTCCA AGATGGAGGA CCCGGGCAGT GTGCTCTCCA CCGCTGTGG
9101 AACTCCGGGA TACGTGGGTG CGGAGGGCCC TGGGCTGGGG CTGTGATGGT
9151 GGGGGGAACC AGGAGTTGAA GGGCAGAGAT TTGTCACCAC CACGTCCTCT
9201 TCCCTCCACA GCCCTGAAG TCCTGGCCCA GAAGCCCTAC AGCAAGGCTG
9251 TGGATTGCTG GTCCATAGGT GTCATCGCCT ACATCTTGTA AGTGGGGCTT
9301 GGCCATGGTA GGCTGTGGCT CCAGAGTTGT CCTCTCGCCT ACTTTCCTCT
9351 CTTCTTCCT CTGCTCTCCC TCTGCCCTCC CTTCTTCCC TCCCTCCCTT
9401 CCTTCCACCA ATCAATTACC AGTATTACTT CATTCAATAG ATACTATGTT
9451 TCAAGCACTG TGCCAAGCAA GCACTGGGGT AAATTTAGCA CAGCACAAC
9501 CAGACAAAGT GCCTGCCCTC AGGGAGCTGA CTTTCTTTCT AGTAGGGAAG
9551 ACAGACAATC AACAAGTAAA TAAATCTACA AACTGACGTC AGGTGATAAA
9601 AATAAATACT GTGGAGAAAA ACCAAGCAGG AATAGGGAGA CGGGGTGATG
9651 CCATTTCACT AGGGAGGTCA GGGAAGGGCT CGCTGTGGAG GTGATGACCG
9701 AGTGGTGAGG GAGCCAGACA TTGGAGGTGT GGGGAAAGAG TGGCATAGGC
9751 AGAAGCAATG GCAAGTGCAA AGGCCCTGAG GAGGGCAAGA TGGCGGCACA
9801 TACAAGGAAC AGAAAGGATA ATGTAGCTAG AACAGGAGTG AGCAGGCAGG
9851 GCTGGTAGAG TTTATAAAGG GGGAATCCT TCCATGGCTC CTGCCTGACC
9901 CCTGAGACTG CCCCAGTGCT CCACCCCGGA GCCAACGGCA CCCGAAAGTG
9951 GAAATGAGGA TGAGTTTCTC CTTGCCCAGG CTCTGCGGTT ACCCTCCCTT
10001 CTATGACGAG AATGATGCCA AACTCTTTGA ACAGATTTTG AAGGCCGAGT
10051 ACGAGTTTGA CTCTCCTTAC TGGGACGACA TCTCTGACTC TGGTATTTGG
10101 GGCTTTGCTT TTTTCCCCTG GGCCCTGCCT CTGGTTCCTC CCTCACCTGC
10151 TTTGGGGGCG GTCTCCCTCC TGCCTTCCTT CTGTCGGATT TTCCAGCACC
10201 ACACAAAGAG CTGTCTTCGA GACCAGACAC CCTACCCCTT CTTCTTCTG
10251 CTTGGGTACT TCCTTCTGCT TGGCTCCCAG AGTGAGAAAC TAGGCATTCA
10301 TTTGTTCAAT CTTCAAACAT AGTCTATTTG AAAATACCTC TCCCTATTG
10351 ACACCTAAT GTCTAAAAACA CCACCATAAA CATTTTCATC CTCCTTTTGT
10401 GCCCCCTATT AAGAAGCAAA CCTGTGAAGC TACTATCGTT TATCATCAGT
10451 GTGAATGCAC TGAGATTAGT CAAGAACAAC TTTTCTTTCT TTTTCTTTCT
10501 TTTTGTGAGC GCAGTCTCGC TCTGTTGCC AGGCTGGAGT GCAGTGGCAC
10551 AATCTCGGCT CACTGCAACC TCTGTCTCCC GGGTTCGAGC AATTCTCTGC
10601 CTCAGCCTCC CAAGTAGCTG GGATTACAGG CGCCACCAC CATGCCCGGC
10651 TAATTTTTTT TGTATTTTTA GTAGAAACAA GGTTCACCA TCTTGGCCAG
10701 GCTGGTCTTG AACTCCTGAC CTCGTGATCC ACCTGCATTG GCCTCCCAAA
10751 GTGCTGGGAT TACAGACATG AGCCACTGTG CCCGGCCATA TGTTTTTCTT

```

FIGURE 3D

10801	AAGAGAGAAA	GGAAAGAGCT	GGAAGGCACG	GGGTGGGAGG	GCCTGAAGAA
10851	GAGCATAGGT	TGGGTGGGGT	GGGGCATGGA	CTGATTTGGC	CTCTTTGTCT
10901	TGATGCCAGG	CCAGACCTGA	GGGAGTGGGT	ATGCTCTTGG	GGAGTACACA
10951	GGCAGTACCA	TGCTGTCATT	ATCTTTGCTT	TTGTCTTGGG	GGTTTAGCCA
11001	AAGATTTTCAT	CCGGCACTTG	ATGGAGAAGG	ACCCAGAGAA	AAGATTCACC
11051	TGTGAGCAGG	CCTTGCAGCA	CCCATGGTGA	GAATTCACAC	AACCTGTGAG
11101	CTGGGGCGGG	ATTTGGGGCC	CTCAGGTCTG	CTTCTGCCCT	CATAGGCAAC
11151	CCACCACATA	ACCCCATCCT	AGGATTGCAG	GAGATACAGC	TCTAGATAAG
11201	AATATCCACC	AGTCGGTGAG	TGAGCAGATC	AAGAAGAACT	TTGCCAAGAG
11251	CAAGTGGAAAG	GTGAGTCCAT	ATCCCTAGTT	CTGGTCCCAG	CCTCCCCCAGG
11301	ACTCCTCCCC	ATCCCTACCC	AGGCTCAGCT	TGCACAGCAC	CTGGCATCAC
11351	ACTGGGCACA	CAGTAACTGC	TTAGGGATCC	TTACTGAAGG	ACTTCATTCA
11401	TTCACTCTTT	CATTCAACAA	ACACTCCCAA	CACCTTCTCT	ATTCCAGAGA
11451	GGGTCCCTCA	CCTCCAAGTC	TAGAGGAAGA	AGTCTGTAAT	TCTTCAGGAG
11501	GCATCTGATC	CAGCCTATGG	GGTCCGAGAA	AGGTCATAAA	AGTGGTGATG
11551	ACCTGACAGA	GCTGTCAGTT	AAGTAGGAAT	TAGTGAGGCA	TAGCGGAATA
11601	ATGTCTATAG	CCATTCCGGG	AAGTGCAAGT	GCTAAGCCTG	GCCAGACTGG
11651	AGGGGCTGAG	GGGACTGAGA	GGCAGGAGCC	CAATTTAGAG	AAGCAGGTAA
11701	GGGGCCAGGC	CTCTTAGGGC	CTCATATGCC	ACAGAGGAGC	ACCAACTTGA
11751	TCCTGAGGGC	ACTGAGGAGC	CCCAGAAGAA	TCTTAGGCAA	GTATTTGCTG
11801	CATAGAAAGG	GCTCTCAGGG	CCAGGCATGG	TGGCTCACGC	CTGTAATCCC
11851	AGCACTTTGG	GAGGCCGAGG	TGGTTGGATC	ACCTGAGGTC	AGGAGTTCAA
11901	GACCATCCTG	GCCAACATGG	CAAAACCCTG	TCTCTACTAA	AAATAAAAGA
11951	ATTAGCCACA	CATGGTGGTG	CGTGCCTGTA	ATCCCAGCTA	CTTGGGAAGC
12001	TGAGGCAGGA	GAATTACTTG	AACCTCGGAG	ATGGGGGTTG	CAGTGAGCTG
12051	AGATCGCGCC	ACTGCACTCC	AGCCTGGGCA	ACAAAGTGAG	ACTCCACCTC
12101	AAAAAAAAAA	GAAAGAGCTC	TCAGGATGCA	GAGAATGGCA	TGGAGTAAAG
12151	ACTGGGTGAC	GCATTAGGAG	GCTGTGGCAG	AGATACAGGC	AGGAGATGGT
12201	AAGGGTTTGG	AACCACAGTA	GCAGCAACAG	GGGGCAGAGA	ACAGTGGTTG
12251	ATCCAGGAGT	CATTTAGGAG	GTGAAACTGA	CAAGACATGA	CGATGCAATG
12301	GATGTTGGGG	GAAAGAGATG	TCAAGGGCTG	GCCCAAGACT	GTGGCTGGGA
12351	ACAGAATGGA	TGGTGGTGGT	ACCATGACTG	AGATGGTTAT	CACAGGGACA
12401	GAAACATGTT	TTGGGGGGAT	GGTTTTAGTT	TTAGACATGG	TGAATTTGAG
12451	GGGTGTGTGG	GACACCTAGG	TGGAGATATT	GAATAGAGAC	ACACCTGAGC
12501	AAGTTACTTC	AGCTTTCTGT	GCCTCAGTTT	CCTCCTTTGA	AAATGATAAT
12551	AGTACCTACC	TCAAAGACTT	TCATGAAGAT	TAAATGAATT	ACTACGTAAA
12601	GTGCTTAGAA	CAGTGCCTGA	CATACAGTGC	TATAGTGTTT	GCTATTACAT
12651	ATTAATATGA	ATTATAGTTA	TGTTTCTATT	TATATATATA	GATACACATA
12701	CATCTAACAT	ATGTGCGTGT	GTGTGTGTAA	ATATATAATA	AAGCCTTGTA
12751	GAGGTTTTTG	GGGGGCTTTA	GGGGAATTAA	TAAAATAACT	CCTGAATGAA
12801	AATAACAGAA	CAATTGCAAG	AATCCCCTG	CGCCCCTGCC	CCATGACTTG
12851	ACTCTCTCAA	AAGTCCTTTC	TCCCCTCTCC	CTTCAATGCC	TTCAATGCCA
12901	GCAAGCCTTC	AATGCCACGG	CTGTGGTGCG	GCACATGAGG	AAACTGCAGC
12951	TGGGCACCAG	CCAGGAGGGG	CAGGGGCAGA	CGGCGAGCCA	TGGGGAGCTG
13001	CTGACACCAG	TGGCTGGGGG	TGAGGAGCGG	GCTCTGCAGA	AGGGCATGGG
13051	TGGTCCACAA	AGGTGCACCC	GGGCTGGAGT	GGAGGGCCTG	CCCCTGCCGC
13101	CACCTCTGTT	CTGTCTTCCC	ATGCAGGGCC	GGCAGCTGGC	TGTTGCTGTC
13151	GAGACTGCTG	CGTGGAGCCG	GGCACAGAAC	TGTCCCCCAC	ACTGCCCCAC
13201	CAGCTCTAGG	GCCCTGGAGG	TCGGGTCTATG	ATCCTCTGCG	TGGGAGGGCT
13251	TGGGGGCAGC	CTGCTCCCCT	TCCCTCCCTG	AACCGGGAGT	TTCTCTGCCC
13301	TGTCCCCCTC	TCACCTGCTT	CCCTACCACT	CCTCACTGCA	TTTTCCATAC
13351	AAATGTTTTCT	ATTTTATTGT	TCCTTCTTGT	AATAAAGGGA	AGATAAAAACC
13401	ATCCTTAGCG	CTGTCTCCCT	CAATATCCCC	CACCCCATCT	TGTTGTGCAA
13451	ACTGACTGCT	TGATTTGGGG	GTGCCTGGCC	TTTGAGGTAG	TCACAGGGAG

FIGURE 3E

Sim4 results:

Exon: 3000-3082, (Transcript Position: 1-83)
 Exon: 7535-7609, (Transcript Position: 84-158)
 Exon: 7696-7834, (Transcript Position: 159-297)
 Exon: 8991-9117, (Transcript Position: 298-424)
 Exon: 9212-9287, (Transcript Position: 425-500)
 Exon: 9980-10092, (Transcript Position: 501-613)
 Exon: 10998-11076, (Transcript Position: 614-692)
 Exon: 11173-11260, (Transcript Position: 693-780)
 Exon: 12902-13019, (Transcript Position: 781-898)
 Exon: 13127-13209, (Transcript Position: 899-981)

CHROMOSOME MAP POSITION:

chromosome 3

ALLELIC VARIANTS (SNPs):

DNA

Position	Major	Minor	Domain
1892	A	G	Intron
3351	G	A	Intron
8636	T	A	Intron
8805	T	C	Intron
9802	T	A	Intron
9833	C	G	Intron
11352	C	T	Intron
13319	T	C A G	Intron
13659	C	G	Intron
14292	G	C	Intron

Context:

DNA

Position

1892 GGTGTTTCGGGTGCAGGGCTGTCTGGGGCACTGTGTGGTGTGGACATGTGTGCTGATGTCT
 GGGAGTACATGTATGATCAGGTGTCACGGGATGTGGATACAAGGCGTACTGGATCTGGGA
 GGCAGGTGTTTGTAGTTTCAGGGCTGTGGAGGGGGCTTGGTGTGGCATGTCTGCTACAGGGA
 TGTGTGTGGATCTGTGAGGGTTGTATTTGGTAGGCCTCCATGTGGGTTTCAGACTCTGCC
 TCTAGAGCTTACACTCGAGTCTCCTTTCTAGAAGATTCTGCCCCCTGGATGGGTGGGCAG
 [A,G]
 GTCCCCCTGGGAAAAAGGTCCTGTTCCAGGAGTGAATCTCACACCAGAGGCCCTAGTCAG
 GGCACCTTCTCCTCATTCTCCCTTAGAGAAAAAGAGAGAAGGAAAGTGCTCTCCCTGAGG
 TCACAAAGCATGCTGGGCTCTGTTTTGGCCTCATCTGTGGATGGGTTGGGAGGCTGTGTT
 CTCTGAATGGGGCCCATTTCTGGCTTCATATTGGAAGTACCAGCCAAGGCCATTTCGATGGC
 CTTTGCCCTCAGCAAGCTTAGCTGGGGGCCCCAGGCCAGGTGTCATTAGGGCCTCTGGAG

3351 GACATCTACGACTTCCGAGATGTTCTGGGCACGTGAGTCCAGGGCAGGATTGGGTGCTGG
 ATGGCTGAGGGAGGCTGAGTCCAGGGTGGGGCTTCCTCTGGTCAATTAATGCTTCCTGTT
 TCCCACAGCCCAGGCCCTGTGGCAGCACTATCTAGGGCCTAAACTGTCCCCAGCTTTTCA
 CTTCTGGATGACAGTGGGTGGGACACGGGCTGCTCTCCCAATAGCCCTGGGTTCTTGAAG
 AGAAAGAAGTCGAGAGAATGAAGGTGCCAGTCAGTCCATTTAACTTGCTGCCAAGAGCTA
 [G,A]
 GTGTTCTAGCCTAGGTTTGGGAACTGAGGCTGGAGATGGCTCTGTTCTTGGTGCTGGGAA

FIGURE 3G

TGCAGAAATAACTCAAACCTGGTCTCTGCCCTTCAAGTTGATCCCAGACATGTGCAAGAG
ACAGACCTACAGAAAATGACAACAGGGTGTGTGCTGTGCTCCAATTAAGGTTGGGATTGA
GGGCTTTGTGGAGCCCAGAGAGAGCTGTGCCTTCTGCCTGGGGGAAAACTTCCTGGAGAA
TGGGGCATTAGAGCTGGGGACTGAAGGATGGGTAGGTGTGCACTTGTGAGAGGAAGAA

8636 AGAGCGAGCCTGGACGACAGAGCGAGACTCCATGTCAAAAAAATAAAATAAAAAACAAA
AATCCTATTCCCCTTCTGTAGAAAACCTTGGATGGGACAGCAAAACATAAAGAAAAAAGCC
AGAAATCCCCGAAATCCTACTCCTCGGAAATAGCGACGGGGCTCACATTTAGCAGTACAT
CTCAATCCGTTCTAGGAGAAGGGCACTTGGGGTGTGACATGCCTGGTTTTGAATTCTGGC
TCTGCTACTGCCTAACTGTGGGTTCTTGGGTGAGTCACTTGCCTCCAAAGGCATCAGTT
[T,A]
CCTCATCTGTTAGGTGAGATTATACAGACTGGCCTAGCAGGGAAGCAGTGAGGATGGCAT
TAAATCAAGCACTAATCCAGGGTCTGGCATAAAATAGGCATTCAAACATTCTTTAGGGC
TTTACAGTGACACCTGAGGTTTAGAGACAGTTCCCCCCCACACCCTCTTGAGCCTTGTC
CTTCTGGAATTTTGGCCTTCTTGAGAGCTTCTTGATTTTCTTATGACAGCCATGAAG
CCACAGTGGCTTTTGGGGATCCATTATTTCTCAGAAGGTGCTTGGAGCGGCAGAAGGTTT

8805 TAGCAGTACATCTCAATCCGTTCTAGGAGAAGGGCACTTGGGGTGTGACATGCCTGGTTT
TGAATTTCTGGCTCTGCTACTGCCTAACTGTGGGTTCTTGGGTGAGTCACTTGCCTCCAA
AGGCATCAGTTTCTCATCTGTTAGGTGAGATTATACAGACTGGCCTAGCAGGGAAGCAG
TGAGGATGGCATTAAATCAAGCACTAATCCAGGGTCTGGCATAAAATAGGCATTCAAACA
TTCTTTAGGGCTTTACAGTGACACCTGAGGTTTAGAGACAGTTCCCCCCCACACCCTC
[T,C]
TGAGCCTTGTCTTCTGGAATTTTGGCCTTCTTGAGAGCTTCTTGATTTTCTTATGA
CAGCCATGAAGCCACAGTGGCTTTTGGGGATCCATTATTTCTCAGAAGGTGCTTGGAGCG
GCAGAAGGTTCTACCAGCCTTAACCATCTCTGATTGCCCTTCTTCCCTCCTGCCCT
TCAAGCCAGAGAATCTGCTGTACTACAGCCTGGATGAAGACTCCAAAATCATGATCTCCG
ACTTTGGCCTCTCCAAGATGGAGGACCCGGGCAGTGTGCTCTCACCGCCTGTGGAAGTC

9802 AGACAAAGTGCCTGCCCTCAGGGAGCTGACTTTCTTTCTAGTAGGGAAGACAGACAATCA
ACAAGTAAATAAATCTACAACTGACGTGAGGTGATAAAAAATAAATACTGTGGAGAAAA
CCAAGCAGGAATAGGGAGACGGGGTGTGCCATTTAGTAGGGAGGTCAGGGAAGGGCTC
GCTGTGGAGGTGATGACCGAGTGGTGGGGAGCCAGACATTGGAGGTGTGGGGAAAGAGT
GGCATAGGCAGAAGCAATGGCAAGTGCAAAGGCCCTGAGGAGGGCAAGATGGCGGCACAT
[T,A]
CAAGGAACAGAAAGGATAATGTAGCTAGAACAGGAGTGAGCAGGCAGGGCTGGTAGAGTT
TATAAAGGGGGAACCTCTTCCATGGCTCCTGCCTGACCCCTGAGACTGCCCCAGTGCTCC
ACCCCGGAGCCAACGGCACCCGAAAGTGGAATGAGGATGAGTTTCTCCCTGCCAGGCT
CTGCGGTTACCTTCCCTTCTATGACGAGAATGATGCCAACTCTTTGAACAGATTTTGAA
GGCCGAGTACGAGTTTACTCTCTTACTGGGACGACATCTCTGACTCTGGTATTTGGGG

9833 TTCTTTCTAGTAGGGAAGACAGACAATCAACAAGTAAATAAATCTACAACTGACGTGAG
GTGATAAAAAATAAATACTGTGGAGAAAAACCAAGCAGGAATAGGGAGACGGGGTGTGCC
ATTTAGTAGGGAGGTCAGGGAAGGGCTCGCTGTGGAGGTGATGACCGAGTGGTGAGGGA
GCCAGACATTGGAGGTGTGGGGAAAGAGTGGCATAGGCAGAAGCAATGGCAAGTGCAAAG
GCCCTGAGGAGGGCAAGATGGCGGCACATAACAAGGAACAGAAAGGATAATGTAGCTAGAA
[C,G]
AGGAGTGAGCAGGCAGGGCTGGTAGAGTTTATAAAGGGGGAACCTCTTCCATGGCTCCTG
CCTGACCCCTGAGACTGCCCCAGTGCTCCACCCCGGAGCCAACGGCACCCGAAAGTGGA
ATGAGGATGAGTTTCTCCCTGCCAGGCTCTGCGGTTACCTTCCCTTCTATGACGAGAAT
GATGCCAACTCTTTGAACAGATTTTGAAGGCCGAGTACGAGTTTACTCTCTTACTGG
GACGACATCTCTGACTCTGGTATTTGGGGCTTTGCTTTTTTCCCTGGGCCCTGCCTCTG

11352 GTGAGCAGGCCTTGCAGCACCCATGGTGAGAATTCACACAACCTGTGAGCTGGGGCGGGA

FIGURE 3H

TTTGGGGCCCTCAGGTCTGCTTCTGCCCTCATAGGCAACCCACCACATAACCCCATCCTA
GGATTGCAGGAGATACAGCTCTAGATAAGAATATCCACCAGTCGGTGAGTGAGCAGATCA
AGAAGAACTTTGCCAAGAGCAAGTGGAAGGTGAGTCCATATCCCTAGTTCTGGTCCCAGC
CTCCCCAGGACTCCTCCCCATCCCTACCCAGGCTCAGCTTGACAGCACCTGGCATCACA
[C,T]
TGGGCACACAGTAACTGCTTAGGGATCCTTACTGAAGGACTTCATTCACTCTTTCA
TTCAACAAACACTCCCAACACCTTCTCTATTCCAGAGAGGGTCCCTCACCTCCAAGTCTA
GAGGAAGAAGTCTGTAATTCTTCAGGAGGCATCTGATCCAGCCTATGGGGTCCGAGAAAG
GTCATAAAAGTGGTGATGACCTGACAGAGCTGTGAGTTAAGTAGGAATTAGTGAGGCATA
GCGGAATAATGTCTATAGCCATTCCGGGAAGTGCAAGTGCTAAGCCTGGCCAGACTGGAG

13319 GGTGAGGAGCGGGCTCTGCAGAAGGGCATGGGTGGTCCACAAAGGTGCACCCGGGCTGGA
GTGGAGGGCCTGCCCCTGCGGCCACCTCTGTTCTGTCTTCCCATGCAGGGCCGGCAGCTG
GCTGTTGCTGTGAGACTGCTGCGTGGAGCCGGGCACAGAACTGTCCCCACACTGCCCC
ACCAGCTCTAGGGCCCTGGACCTCGGGTCATGATCCTCTGCGTGGGAGGGCTTGGGGGCA
GCCTGCTCCCTTCCCTCCCTGAACCGGGAGTTTCTTGCCCTGTCCCCTCCTCACCTGC
[T,C,A,G]
TCCCTACCACTCCTCACTGCATTTTCCATACAAATGTTTCTATTTTATTGTTCTTCTTG
TAATAAAGGGAAGATAAAACCATCCTTAGCGCTGTCTCCCTCAATATCCCCACCCCATC
TTGTTGTGCAAACTGACTGCTTGATTTGGGGGTGCCTGGCCTTTGAGGTAGTCACAGGGA
GGCCCCCTCCCCAACATGAGACTGGGTGGGGATGGGGAGAGAGAAGTGGGGAATGGAGGGG
AAGTGCTTGGGGAATTTCTTTGTCCAGGGTGCCCCATCTAGCCTTCCGGCCCTTTGGAA

13659 CTATTTTATTGTTCTTCTTGTAAATAAAGGGAAGATAAAACCATCCTTAGCGCTGTCTCC
CTCAATATCCCCACCCCATCTTGTTGTGCAAACTGACTGCTTGATTTGGGGGTGCCTGG
CCTTTGAGGTAGTCACAGGGAGGCCCCCTCCCAACATGAGACTGGGTGGGGATGGGGAGA
GAGAAGTGGGGAATGGAGGGGAAGGTGCTTGGGGAATTTCTTTGTCCAGGGTGCCCCATC
TAGCCTTCCGGCCCTTTGGAACCCCTTCTGCGCTTTGCTGGTGGCTCCTGAGCATGGCGG
[C,G]
ATTGGCGCAGGTCGGCACTGAACAGCACCTGTAGGAGGGTGGAGTCTGTGTGGGGAGGAG
GGTACACTGGGGTCAGGGCTGGTGAGACTAGTGACAGTGTGGGAGGTGGAAGAGTCCTT
GGGGAACAGGGCCGAAGGCAATGAGAATCCACTGGGGTTGGGACAGGGGTGGCTGGAGAG
TCCTTTAGGGCCACCTGGGGCGGTGGTGGAAGAGTCCACTGGGTCTGGGCTGGAGGAGAG
GAAACCTAGGGAGGACACCTAGGTACACTCACCGCTTGGGCCAGCCAGCATAAGGTCCC

14292 AGTGATGAGACTAAGTTATCTGACCCCTTCTGTGACCCATCAACAGAAGTAGGGTCTGAG
GGAGAGGTGACTAAGAGAGAGAGAAGTTTCTACCATCCCAGCCCACTGCCAGCCCCTGCA
GCCACTTTCTCACCAGTTCTTGTGTTGGTCTGGGGGCTCGTCCCTTCGCCTGGGACG
TGGTAGGGTGCCAGCTGTAGTCACGTTGGGCAATGTGCCACATATGGACATCCACGGGCA
CAGCCTGGGGCTTGTCTAGGGCCATCAGGCAGATGCAGTCAGCCACCTTTGACAGACACA
[G,C]
AATGAGCCCTTGTGGAAGAAGGGCAGCATGTGGCCAGCATCTTGCTTATAGCCCCAAAGC
CGGCTGCTTTCTCCTTCACTCTGGGGTTACTGTTGTTCTATATTCTCAATCAACAGATAC
TATCTATGAATACACTTTTTTTTTGTTTGTGTTTGGATGGAGTCTCGCTCTGTTGCCTA
GGCTGGAGTGCACTGGTGCAATCCTGGCTCTCCAGGTTCAAGCAATTCTCCTACCTCAG
CCTCCCAAGTAGCTGGGATTACAGGCATGTGCCACCACGTGTGGCTAATTTTTGTGTTTT

FIGURE 3I